

# Human Time-Frequency Acuity Beats the Fourier Uncertainty Principle

Jacob N. Oppenheim and Marcelo O. Magnasco\*

*Laboratory of Mathematical Physics, Rockefeller University, New York, New York 10065*

(Dated: August 24, 2012)

The time-frequency uncertainty principle states that the product of the temporal and frequency extents of a signal cannot be smaller than  $1/(4\pi)$ . We study human ability to simultaneously judge the frequency and the timing of a sound. Our subjects often exceeded the uncertainty limit, sometimes by more than tenfold, mostly through remarkable timing acuity. Our results establish a lower bound for the nonlinearity and complexity of the algorithms employed by our brains in parsing transient sounds, rule out simple “linear filter” models of early auditory processing, and highlight timing acuity as a central feature in auditory object processing.

PACS numbers:

Fourier transformation turns signals “inside out”, in the sense that low frequencies dictate what happens at long times, while high frequencies are needed to create fine temporal detail. A specific instance of this general property is Fourier’s uncertainty theorem, which states that considering the absolute value squared of a signal  $x(t)$  as a probability distribution in time,

$$P(t) = \frac{|x(t)|^2}{\int_{-\infty}^{\infty} |x(t')|^2 dt'} \quad (1)$$

and the absolute value squared of its Fourier transform  $\tilde{x}(f)$  as a distribution in frequency,

$$P(f) = \frac{|\tilde{x}(f)|^2}{\int_{-\infty}^{\infty} |\tilde{x}(f')|^2 df'} \quad (2)$$

then the product of the standard deviations

$$\Delta t = \sqrt{\text{var}(t)} \quad \text{and} \quad \Delta f = \sqrt{\text{var}(f)} \quad (3)$$

is bounded from below [1]:

$$\Delta t \Delta f \geq \frac{1}{4\pi} \quad (4)$$

whence it is inferred that signals with a small temporal extent require many frequencies for their representation.

The theorem refers to the original signal and its Fourier transform. In time-frequency analysis one attempts to describe a signal in the two-dimensional time-frequency plane, akin to a musical score which has time as its horizontal axis and frequency as its vertical axis. Here the uncertainty principle begets the Gabor limit [1, 2]. In this setting, the emphasis is shifted from the uncertainties being a property of the signal, to their being a property of the transform itself, leading to an important distinction between resolution and precision. Resolution refers to our ability to verify that two objects are distinct, while precision refers to our ability to track the parameters of a single object, given prior knowledge or assurance it is only one component. This distinction is well-established in optics, where it is known the

wavelength of light limits resolution, so two glass beads can not be resolved as being different if they are closer together than a wavelength, yet does not limit precision, since a single bead can be tracked with nanometer accuracy. In time-frequency analysis, it has been proven that first-order operators cannot exceed the uncertainty bound [2]. However, there are many nonlinear distributions that can achieve higher precision than the Gabor limit when applied to isolated signal components; these include quadratic (Cohen’s class) distributions like Wigner-Ville [3] and Choi-Williams [4], and many higher-order ones, such as multi-tapered spectral estimates [5, 6], the Hilbert-Huang distribution [7], and the reassigned spectrograms [8–12]. All such distributions achieve high precision, yet give interfering results when two signals are closer together than an uncertainty envelope. Our experimental test is designed to directly measure precision, not resolution.

A key goal in neuroscience is to establish which algorithms the brain uses to process perceptual information. Psychophysics, by establishing tight bounds on the performance of our sensory systems, is sometimes able to rule out entire families of perceptual algorithms as candidates when they, as a matter of principle, cannot achieve the expected performance [13, 14]. We shall show below that human subjects can discriminate better, and occasionally much better, than the uncertainty bounds. This categorically rules out any first order operators, such as the standard sonogram, from consideration, and puts a stringent bound on the performance of any candidate algorithm. This is relevant both in the scientific and technical areas (e.g. [15]), as many high-level models of auditory processing assume an underlying representation of the earliest steps in auditory information homologous to a bank of linear filters [16, 17]. In many applications such as speech recognition or audio compression (e.g. MP3 [18]), the first computational stage consists of generating from the source sound small sonogram snippets, which become the input to latter stages. Our data suggest this is not a faithful description of early steps in auditory transduction and processing, which appear to preserve

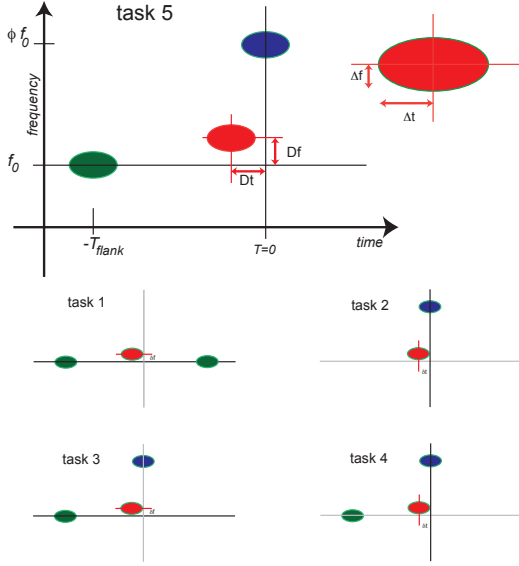


FIG. 1: (A) Stimulus and task. In our final task 5, subjects are asked to discriminate simultaneously whether the test note (red) is higher or lower in frequency than the leading note (green), and whether the test note appears before or after the flanking high note (blue). For each instance of the task, two numbers are generated ( $Dt$  and  $Df$ ) and two Boolean responses (left/right, up/down) are recorded. Tasks 1 through 4 lead to this final task: task 1 is frequency only (uses two flanking notes), task 2 timing only, task 3 is frequency only but with the flanking high note (blue) as a distractor, and task 4 is timing only, with the leading (green) note as a distractor. (B) Measurement strategy. As task 5 proceeds, the numbers  $Df$  and  $Dt$  are drawn as Gaussian random numbers with variances  $pref$  and  $multf$ . The smaller these variance, the harder the task. The variances are independently controlled by a 2D1U (two down, one up) procedure: when two responses in a row are correct, the variance is reduced and the task is made harder; the variance is increased 0 for every wrong response. This procedure converges to a demanding regime, one in which the subject makes frequent mistakes, but fewer than 50%. Data shown from subject **qr3zb**[21]. (C) **Datum definition.** We show in red the time responses of subject **qr3zb**; horizontal axis is  $Dt$ , vertical axis is 0 (for before) or 1 (for after); we have slightly offset the data by random amounts from 0 or 1 to be able to visualize the density of points at any given  $Dt$ . In blue, the psychometric curve which maximizes the likelihood of the data. The procedure described in 1(b) has converged to a high density of tests around 0, spanning with high density the area in which the psychometric curve most rapidly rises[21].

much more accurate information about the timing and phase of sound components than about their intensity [12, 19, 20].

We shall carefully distinguish between the physical attributes of the stimulus and the homologous psychological quantities. Most relevant will be the distinction between  $\Delta t$  and  $\Delta f$ , the *physical* uncertainties defined by

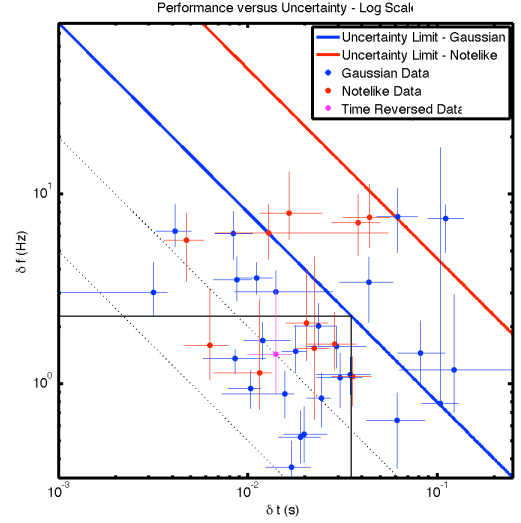


FIG. 2: Figure 2. Summary of main results: discrimination limens for each test. Each round dot is a completion of Task 5 by a subject on an individual day, with at least 100 presentations. There were 12 subjects totaling 26 individual sessions for Gaussian and 12 sessions for notelike tests. Blue denotes Gaussian packet while red denotes notelike. The two solid lines are the locus of the relation  $\delta t \delta f = \Delta t \Delta f$ ; any dots below these curves violate the corresponding uncertainty relation. Error bars in both dimensions were obtained by generating 1000 bootstraps from the raw data and plotting the 25% – 75% quartiles. Raw data provided in Suppl. Table S1.

eqns (1, 2), versus  $\delta t$  and  $\delta f$ , the *psychological* limens of discrimination in time and frequency. It would be absolutely trivial to violate the theorem by using an incorrect definition of the  $\delta t$  and  $\delta f$  or an incorrect evaluation of the bound. Therefore the limens should be carefully defined to carry the equivalent meaning of a standard deviation, so that the actual number is directly comparable to the equivalent physical attribute. It is standard in the literature to operatively define limens of discrimination through a same-different paradigm. For many reasons that are detailed below, but particularly because same-different is unlike the standard deviation definition of the physical  $\Delta t$  and  $\Delta f$ , we shall operatively define  $\delta t$  and  $\delta f$  through a two-alternative forced choice above or below paradigm, and then regress by maximum likelihood the performance data against a psychometric curve in the form of an error function; the standard deviation parameter of this error function is our limen.

We test for both limens simultaneously as a single task. Prior work in the area (e.g.[22–24]) has always concentrated on comparing measurements of frequency discrimination limens  $\delta f$  against the physical temporal duration of the sound packets'  $\Delta t$ . This is inadequate for our purposes on two grounds: first that it treats both quantities in the inequality differently, contradicting the "spirit" of the uncertainty principle, and second because it fails to

verify human ability in an important and ecologically relevant domain, namely timing acuity.

We study simultaneous time-frequency acuity for two test stimuli [21]. The first one is a Gaussian packet, for which  $4\pi\Delta t\Delta f = 1$ , attaining the bound in the theorem; our study shows that many subjects display limens such that  $4\pi\delta t\delta f \ll 1$ . In most of the subjects, the overall increase in performance comes from substantial increases in timing accuracy. One of our subjects, ar4tl, for example, when tested with notes of  $\Delta t = 35ms$  attained a limen of  $\delta t = 3ms$ , while the frequency performance degraded,  $\delta f > \Delta f$ . In our second test we study perception of a wave packet with a note-like envelope characterized by a rapid rise and a slow exponential decay. Such envelopes are sub-optimal according to the uncertainty principle, having a product  $4\pi\Delta t\Delta f$  strictly bigger than one; in our case,  $4\pi\Delta t\Delta f = 5.7079$ . However the performance of our subjects on such packets is just as good, if not better, than their performance on the Gaussian packet; this has broad implications for understanding early auditory processing.

The results from task 5 are summarized in Figure 2. Each dot there corresponds to a simultaneous limen measurement as outlined above. Some subjects performed several different measurements, never on the same week. Two extremes are worth discussing in detail. The lowest blue dot at the bottom center of the plot displays the greatest violation of the principle in our records, by a factor of about 13. The subject qr3zb displayed in equal measure a marked increase in frequency acuity as well as temporal acuity, and hence the measurement is below and to the left of the physical values of  $\Delta t$ ,  $\Delta f$  for the Gaussian note (indicated by the black lines). The subject is a professional musician and possesses absolute pitch. The second point to consider is the leftmost point, at the center left of the diagram, from subject ar4tl. This is the smallest  $\delta t$  limen in our records; at 3 ms, the subject was able to discriminate the relative timing of two notes by a factor of 13 better than their widths; it should be noted that 3ms is barely more than a single period of the test note, 2.27ms. However this subject was unable to estimate frequency better than its physical extent, which is indicated by the dot being above the black line indicating the Gaussian  $\Delta f$ , so overall this measurement beats the uncertainty principle only by a factor of 10. The subject is an electronic musician who microcomposes and works in precision sound editing.

We can now examine some implications of this data. First, even though the notelike packet's uncertainty product is substantially above the minimum, the subjects seem to be able to discriminate with it just as well as with the Gaussian packet, leading to two measurements (red dot at the bottom of the graph and red dot on the black horizontal line) that beat relative uncertainty by a factor of 50:  $\delta t\delta f \approx (1/50)\Delta t\Delta f$ , and absolute uncertainty by a factor of 10:  $4\pi\delta t\delta f \approx (1/10)$ . Therefore we

may conclude that the larger uncertainty product of the test note does not affect the subjects' acuity, at least in this particular case. Second, the plot shows a number of different strategies that subjects use to discriminate, with a remarkable spread: from those who do not achieve the physical limits in either dimension (1), those who have better frequency but worse timing (4), those with better timing and worse frequency (10), and those who have both better timing as well as better frequency discrimination than the physical values (8). While the relative number of measurements in each category undoubtedly reveals the underlying bias of our subject population, the fact that there are many strategies should be robust. However, even so there is a noticeable shift of the cloud to the left of the reference notes, so that we can see that on median the subjects perform twice as well in timing discrimination as the physical value; 80% of the Gaussian data and 100% of the notelike data lies on the  $\delta t < \Delta t$  halfplane.

It is important to stress the difficulty of the task. Preliminary testing on a handful of non-musicians and amateur musicians yielded no subjects that could break the Uncertainty Principle. Non-musicians found task 3 (frequency discrimination with distractors) difficult, whereas musicians found it trivial, likely due to their practice playing instruments (or singing) when surrounded by other instrumentalists. The ability to ignore, or otherwise compartmentalize out distractors is believed to be one of the major sources of the better uncertainty product results for musicians in task 5. We found that composers and conductors achieved the best results in task 5, consistently beating the uncertainty principle, often by a factor of 2 or more, whereas performers were more likely to beat it only by a few percentage points. After debriefing experimentees, it appears that the necessity of hearing multi-voiced music (both in frequency and in time) in one's head and coaching others to perform it led to the improved performance of conductors and composers.

Early last century a number of auditory phenomena, such as residue pitch and missing fundamentals, started to indicate that the traditional view of the hearing process as a form of spectral analysis had to be revised. In 1951 Licklider [25] set the foundation for the temporal theories of pitch perception, in which the detailed pattern of action potentials in the auditory nerve is used [26], as opposed to spectral or place theories, in which the overall amplitude of the activity pattern is evaluated without detailed access to phase information. The groundbreaking work of Ronken [22] and Moore [23] found violations of uncertainty-like products and argued for them to be evidence in favor of temporal models. However this line of work was hampered fourfold, by lack of the formal foundation in time-frequency distributions we have today, by concentrating on frequency discrimination alone, by technical difficulties in the generation of the stimuli, and not

the least by lack of understanding of cochlear dynamics, since at the time the active cochlear processes had not yet been discovered. Perhaps because of these reasons this groundbreaking work did not percolate into the community at large, and as a result most sound analysis and sound processing tools today continue using models based on spectral theories. We believe it is time to revisit this issue.

We have conducted the first direct psychoacoustical test of the Fourier uncertainty principle in human hearing, by measuring subject performance on a simultaneous temporal and frequency discrimination task. We have meticulously defined our discrimination limens  $\delta t, \delta f$  to correspond to standard deviations of a Gaussian distribution, so as to be directly comparable to the physical attributes  $\Delta t, \Delta f$  as they are defined in the theorem, to avoid the possibility of a fictitious violation of the theorem's bound by usage of incorrect units. We have used as our two paradigmatic stimuli the Gaussian envelope, which is the optimal envelope according to the uncertainty theorem, and a notelike envelope with the sharp onset and shallower decay characteristic of many natural sounds [27] and of cochlear responses.

Our data indicate that human subjects often beat the bound prescribed by the uncertainty theorem, sometimes by factors in excess of 10. This is sometimes accomplished by an increase in frequency acuity, but by and large it is temporal acuity that is increased and largely responsible for these gains. Our data further indicate subject acuity is just as good for a note-like amplitude envelope as for the Gaussian, even though theoretically the uncertainty product is increased for such waveforms. Our study directly rules out many of the simpler models of early auditory processing, often used as input to the higher-order stages in models of higher auditory function. We hope our study will direct further inquiry into which family of time-frequency analysis (e.g. [4, 6, 9, 10, 12]) may be the mechanism underlying the auditory hyperacuity displayed by our subjects. Elucidation of such mechanisms is likely to have wide-ranging applications, both in fields where matching human performance is an issue, such as speech recognition, as well as those more removed, such as radar, sonar and radio astronomy.

We wish to thank Mayte Suarez-Farinas and Maurizio Pellegrino for their algorithmic and psychophysical expertise, Tim Gardner for valuable discussions, and Josh Oppenheim for physiological background. All fitted data are available in the Supplementary Information. Supported in part by NSF grant EF-0928723.

---

\* magnasco@rockefeller.edu

- [1] D. Gabor, *Nature* **159**, 591 (1947).
- [2] L. Cohen, *Time-frequency analysis* (Prentice Hall PTR,

- Englewood Cliffs, N.J., 1995).
- [3] E. P. Wigner, *Physical Review* **40**, 749 (1932).
- [4] H. I. Choi and W. J. Williams, *Ieee Transactions on Acoustics Speech and Signal Processing* **37**, 862 (1989).
- [5] D. J. Thomson, *Proceedings of the Ieee* **70**, 1055 (1982).
- [6] O. Tchernichovski, F. Nottebohm, C. E. Ho, B. Pesaran, and P. P. Mitra, *Animal Behaviour* **59**, 1167 (2000).
- [7] N. E. Huang, Z. Shen, S. R. Long, M. L. C. Wu, H. H. Shih, Q. N. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* **454**, 903 (1998).
- [8] K. Kodera, R. Gendrin, and C. D. Villedary, *Ieee Transactions on Acoustics Speech and Signal Processing* **26**, 64 (1978).
- [9] F. Auger and P. Flandrin, *Icassp-94 - Proceedings*, Vol 4 pp. 317–320 (1994).
- [10] E. ChassandeMottin, I. Daubechies, F. Auger, and P. Flandrin, *Ieee Signal Processing Letters* **4**, 293 (1997).
- [11] S. A. Fulop and K. Fitz, *Journal of the Acoustical Society of America* **119**, 360 (2006).
- [12] T. J. Gardner and M. O. Magnasco, *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6094 (2006).
- [13] H. Fastl and E. Zwicker, *Psychoacoustics : facts and models*, Springer series in information sciences, (Springer, Berlin ; New York, 2007), 3rd ed.
- [14] G. A. Gescheider, *Psychophysics : the fundamentals* (L. Erlbaum Associates, Mahwah, N.J., 1997), 3rd ed.
- [15] F. Le Chevalier, *Principles of radar and sonar signal processing*, Artech House radar library (Artech House, Boston, 2002).
- [16] R. D. Patterson, K. Robinson, J. Holdsworth, D. Mckewon, C. Zhang, M. Allerhand, Decheveigne, and G. Langner, *Auditory Physiology and Perception* **83**, 429 (1992).
- [17] V. Hohmann, *Acta Acustica United with Acustica* **88**, 433 (2002).
- [18] M. Bosi and R. E. Goldberg, *Introduction to digital audio coding and standards*, The Kluwer international series in engineering and computer science (Kluwer Academic Publishers, Boston, 2003).
- [19] E. Covey and J. H. Casseday, *Annual Review of Physiology* **61**, 457 (1999).
- [20] M. Elhilali, J. B. Fritz, D. J. Klein, J. Z. Simon, and S. A. Shamma, *Journal of Neuroscience* **24**, 1159 (2004).
- [21] See Supplemental Material at [URL will be inserted by publisher] for testing procedures and parameters, fitted data, controls, and discussion of performance at other parameter values.
- [22] D. A. Ronken, *Journal of the Acoustical Society of America* **49**, 1232 (1971).
- [23] B. C. J. Moore, *Journal of the Acoustical Society of America* **54**, 610 (1973).
- [24] D. A. Nelson, M. E. Stanton, and R. L. Freyman, *Journal of the Acoustical Society of America* **73**, 2117 (1983).
- [25] J. C. R. Licklider, *Journal of the Acoustical Society of America* **23**, 147 (1951).
- [26] A. de Cheveigné, *Pitch perception models* (Springer, New York, 2005), pp. xvi, 364 p. ill. (some col.) 24 cm., Springer handbook of auditory research v 24.
- [27] E. C. Smith and M. S. Lewicki, *Nature* **439**, 978 (2006).